

Network Working Group  
Internet Draft  
Category: Standards Track

Fatai Zhang  
Suresh B R  
SenthilKumarS  
Jun Sun  
Huawei

Expires: July 21, 2010

January 21, 2010

## UDP as Transport Protocol for PCECP

[draft-zhang-pce-udp-for-pcecp-00.txt](#)

### Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Copyright (c) <2010> IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on July 20, 2010.

[draft-zhang-pce-udp-for-pcecp-00.txt](#)

January 2010

## Abstract

The Path Computation Element Communication Protocol defines a request/response protocol used for the communication between a Path Computation Client (PCC) and a Path Computation Element (PCE), or between two PCEs. PCEP employs Transmission Control Protocol (TCP) as the transport layer protocol by using a registered TCP port.

This document proposes the possibility of employing UDP as the transport layer protocol instead of TCP.

UDP is a connectionless protocol within TCP/IP protocol suite that corresponds to the transport layer in the ISO/OSI reference model. UDP converts data messages generated by an application into packets to be sent through IP. The reliability is not guaranteed by UDP and should be ensured by the application that generates the data message.

The PCECP application is flexible to ensure the reliability of PCECP messages. This document explains on how the reliability of the PCECP messages can be achieved by means of PCECP, while operating PCECP over UDP.

## Table of Contents

<a href="#">1.</a>	<a href="#">Introduction.....</a>	<a href="#">3</a>
<a href="#">1.1.</a>	<a href="#">Requirements Language.....</a>	<a href="#">3</a>
<a href="#">2.</a>	<a href="#">Terminology.....</a>	<a href="#">3</a>
<a href="#">3.</a>	<a href="#">Requirements.....</a>	<a href="#">4</a>
<a href="#">3.1.</a>	<a href="#">Motivations for Using UDP as the transport protocol for PCECP.....</a>	<a href="#">5</a>
<a href="#">3.2.</a>	<a href="#">Benefits from PCECP over UDP.....</a>	<a href="#">6</a>
<a href="#">4.</a>	<a href="#">Proposing UDP for PCECP Transport Protocol.....</a>	<a href="#">6</a>
<a href="#">4.1.</a>	<a href="#">Reliability of PCECP Messages.....</a>	<a href="#">7</a>
<a href="#">4.1.1.</a>	<a href="#">Fast Request Retransmission with Exponential or Linear Back-off Mechanism.....</a>	<a href="#">7</a>
<a href="#">4.1.2.</a>	<a href="#">Retransmission Parameters.....</a>	<a href="#">8</a>
<a href="#">4.2.</a>	<a href="#">Handling Duplication.....</a>	<a href="#">8</a>
<a href="#">4.3.</a>	<a href="#">Congestion Control.....</a>	<a href="#">9</a>
<a href="#">4.4.</a>	<a href="#">PCECP Messages and Object Formats.....</a>	<a href="#">9</a>
<a href="#">5.</a>	<a href="#">UDP Port.....</a>	<a href="#">10</a>
<a href="#">6.</a>	<a href="#">Security Considerations.....</a>	<a href="#">10</a>

<a href="#">6.1. Authentication of PCECP Messages.....</a>	<a href="#">10</a>
<a href="#">6.1.1. RDM Monotonic Counter TLV (64-bits).....</a>	<a href="#">11</a>
<a href="#">6.1.2. HMAC-MD5 TLV.....</a>	<a href="#">11</a>
<a href="#">6.1.3. Message Validation.....</a>	<a href="#">12</a>

<a href="#">6.1.4. Key Utilization.....</a>	<a href="#">13</a>
<a href="#">7. Conclusion.....</a>	<a href="#">13</a>
<a href="#">8. IANA Considerations.....</a>	<a href="#">13</a>
<a href="#">8.1. UDP Port.....</a>	<a href="#">13</a>
<a href="#">8.2. PCECP Objects.....</a>	<a href="#">13</a>
<a href="#">8.3. PCECP TLV Type Indicators.....</a>	<a href="#">13</a>
<a href="#">9. Acknowledgments.....</a>	<a href="#">14</a>
<a href="#">10. References.....</a>	<a href="#">14</a>
<a href="#">10.1. Normative References.....</a>	<a href="#">14</a>
<a href="#">10.2. Informative References.....</a>	<a href="#">14</a>
<a href="#">11. Authors' Addresses.....</a>	<a href="#">14</a>

## [1. Introduction](#)

[RFC4655] describes the motivation and architecture for a Path Computation Element (PCE) based model for the computation of Multi-Protocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs). The PCE is an entity (component, application or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints. The Path Computation Client (PCC) is any client application requesting a path computation to be performed by a PCE. [RFC5440] specifies the Path Computation Element communication Protocol (PCEP) used for communication between a PCC and PCE in compliance with [RFC4657]. The PCECP (PCE Communication Protocol) is an application layer communication protocol employed between PCC and PCE, as well between PCE and PCE (In this document, we refer to all communications as PCC-PCE regardless of whether they are PCC-PCE or PCE-PCE.). The purpose of PCECP is to carry TE-LSP computation requests, responses and other notifications between PCC and PCE or between PCE and PCE. The PCECP operates over a transport layer protocol for transmitting the PCECP messages between the PCECP peers.

[RFC5440] has defined TCP as the transport protocol for PCECP. This document explains how to operate PCECP over UDP instead of TCP, and defines the necessary changes and extensions to PCEP.

### [1.1. Requirements Language](#)

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119](#) [[RFC2119](#)].

## [2. Terminology](#)

The following terminology is used in this document.

zhang

Expires July 2010

[Page 3]

---

[draft-zhang-pce-udp-for-pcecp-00.txt](#)

January 2010

GMPLS: Generalized Multi-Protocol Label Switching

IP: Internet Protocol. IP governs the break up of data messages into packets, the routing of the packets from sender to destination network and station, and the reassembly of the packets into the original data messages at the destination.

PCC: Path Computation Client. Any client application requesting a path computation to be performed by the Path Computation Element.

PCE: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

PCECP: Path Computation Element Communication Protocol, which is a request/response protocol used for the communication between a PCC and a PCE, or between two PCEs

PCEP: A kind of specific protocol defined in [[RFC5440](#)] for PCECP.

PCECP Peer: An element involved in a PCECP session (For example, a PCC or a PCE).

PCECP Session: The PCECP session is a logical connection established automatically between the PCECP peers.

TE LSP: Traffic Engineering MPLS Label Switched Path.

TCP: Transmission Control Protocol. TCP is a connection-oriented, end-to-end reliable protocol that governs the break up of data messages into packets to be sent through IP (Internet Protocol), and the reassembly and verification of the complete messages from packets received by IP.

UDP: User Datagram Protocol. UDP is a connectionless protocol that converts data messages generated by an application into packets to be sent through IP.

### 3. Requirements

In GMPLS networks, the service has to be restored within the minimum restoring time to avoid service interruption in the event of fiber or node failure. In such scenarios, establishing a PCECP session, controlling request retransmission, message/packet overhead over the bandwidth constrained in-band signaling channels becomes critical.

zhang

Expires July 2010

[Page 4]

---

[draft-zhang-pce-udp-for-pcecp-00.txt](#)

January 2010

#### 3.1. Motivations for Using UDP as the transport protocol for PCECP

The following are the motivations for using UDP as the transport protocol:

**Absence of Initial Connection Establishment:** TCP employs three-way handshake mechanism before sending the PCEP messages. Unlike TCP, UDP does not establish end-to-end connection between PCECP peers. UDP communication consequently does not incur connection establishment and teardown overheads.

**Absence of Connection State:** TCP maintains the connection state in the transport layer of PCECP peers. This includes the connection state, receive and send buffers, congestion control parameters, and sequence numbers. This information is used to realize the reliable data transfer of TCP and the congestion control that leads to high memory and CPU usage.

UDP does not maintain any such connection state and does not track the congestion parameters or the sequence numbers. As a result, a PCE can typically support more number of PCCs when the PCECP runs over UDP rather than TCP.

**Minimum Segment Overhead:** The TCP segment has 20 bytes of header overhead in each segment to guarantee the reliable transfer of messages. The UDP only has 8 bytes of overhead.

**Absence of Inherent Data Rate:** TCP has a congestion control mechanism that restricts the PCECP peer when one or more links

between sender and receiver becomes excessively congested. This restriction can have a severe impact for path request/response when a TE-LSP is rerouted. On the other hand, using UDP the application can send data that is constrained by the rate at which the application generates data and the process capabilities of the source.

**Application Flexibility for Reliability:** The built-in reliability mechanism in TCP leads to additional overhead and resource usage in the network. The PCECP can realize its own reliable mechanism to ensure the reliability of messages. For example, the PCC can realize appropriate retransmission strategy to guarantee the reliable delivery of PCECP messages. The PCC can decide to retransmit the request to the same PCE or to a different PCE (based on the availability) in case if no response is received.

**Absence of Message Boundary:** TCP is a stream oriented protocol and a read from the socket does not guarantee the complete message. To receive a complete message, the state oriented message assembler is

required. Unlike TCP, the message assembler is not required for UDP. A single UDP datagram consists of complete message and the message boundaries are guaranteed irrespective of link/path MTU.

### 3.2. Benefits from PCECP over UDP

The following are the advantages of UDP:

- o No time is spent in establishing a TCP connection to the PCE server. Directly PCECP session establishment can be initiated using UDP.
- o Complexity of PCECP message parser is reduced as complete PCECP message is carried by one UDP datagram. This will enhance the PCECP parser performance.
- o PCC has the complete control about the retransmission and can choose a different PCE upon retransmission failure. Custom retransmission algorithm and policy can be implemented based on the network.
- o If the number of PCC is huge, then number of sessions to be handled is also huge. Using TCP needs more memory and more processing. Using UDP this problem can be solved.

By operating PCECP over UDP, the above mentioned requirements are satisfied.

#### 4. Proposing UDP for PCECP Transport Protocol

UDP is a connectionless protocol that provides a minimal, best-effort, message-passing transport with minimum protocol mechanism. The service provided by UDP is unreliable that provides no guarantee for delivery. The simplicity of UDP reduces the overhead from using the protocol and the services may be adequate in many cases.

When PCECP uses UDP as transport protocol, an UDP datagram must contain exactly one PCECP message. So while using UDP, the application has to realize its own mechanisms to guarantee the reliable delivery of messages, handle duplication, and congestion of messages.

The following sections explain on how PCECP can handle the reliable delivery of PCECP messages avoiding duplication and congestion on a UDP based PCECP session.

##### 4.1. Reliability of PCECP Messages

The PCC or PCE originating the PCECP message is responsible for the reliable delivery of PCECP messages. For example, when a PCC is sending a TE-LSP path request to PCE, PCC is responsible to keep track of the request. The PCC must retransmit its PCECP message, if it fails to receive the response message, either the path response or error response from the PCE. [[draft-ietf-pce-monitoring](#)]

###### 4.1.1. Fast Request Retransmission with Exponential or Linear Back-off Mechanism

The PCC will transmit a request message to the selected PCE. The message exchange terminates when either the PCC receives the appropriate PCECP response successfully, or when the message exchange has failed as per the retransmission mechanism.

The retransmission behaviour is controlled and described by the following variables:

RT: Retransmission timeout

IRT: Initial retransmission time

MRT: Maximum retransmission time

MRC: Maximum retransmission count

MRD: Maximum retransmission duration

RAND: Randomization factor. The RAND is a random number chosen with a uniform distribution between -0.3 and +0.3.

The PCC controls the retransmission behaviour using exponential back-off mechanism as explained below.

- o With each request transmission or retransmission, the PCC sets RT. The RT for the first message retransmission is based on IRT.

$$RT = (1 + RAND) * IRT$$

- o On the expiry of RT, if the PCC has not received a response to the PCReq, the PCC recomputes RT and retransmits the message. Each of the computations of a new RT includes a RAND. The RAND is included to minimize synchronization of messages transmitted by PCCs. The RT for each subsequent message transmission is based on the previous value of RT.

$RT = (Back-off * RT_{prev}) + (RAND * RT_{prev})$ , where Back-off = 1 for Linear and 2 for Exponential.

- o The MRT specifies a boundary value for RT (disregarding the randomization added by the use of RAND). If MRT is 0, then there is no upper limit on the value of RT. Otherwise, when  $RT > MRT$ , RT is recalculated using,  $RT = (1 + RAND) * MRT$ .
- o The MRC specifies a boundary value on the number of times a PCC may retransmit a message. Unless MRC is zero, the message exchange fails once the PCC has retransmitted the message MRC times.
- o MRD specifies a boundary value on the length of time a PCC may



retransmit a message. Unless MRD is zero, the message exchange fails once MRD seconds have elapsed since the PCC first transmitted the message.

- o When MRC and MRD are non-zero, the message exchange fails whenever either of the conditions specified in the previous points is met. If both MRC and MRD are zero, the PCC continues to transmit the message until it receives a response, but setting both MRC and MRD to zero is not recommended.

The same retransmission mechanism either exponential back-off or linear mechanism is followed during PCE switching.

#### 4.1.2. Retransmission Parameters

This section presents the default values used to describe the message transmission behaviour.

Parameter	Default Value	Range
IRT	1 sec	0.1 sec to 8 secs
MRC	3	0 to 8
MRT	2 secs	0.5 sec to 16 secs
MRD	8 secs	1.0 sec to 64 secs
MPC	2	0 to 16

#### [4.2. Handling Duplication](#)

UDP does not protect against any message duplication, but PCECP can follow the below mentioned mechanism to gracefully handle duplication. A duplicate PCECP message is identified by a repeated PCECP request-ID received from the same PCC.

- o When a PCE receives a duplicate request message, and if it is still processing the original message (path computation is still in progress or queued, and the path response is not sent), the newly received request message is dropped.
- o When a PCE receives a duplicate request message for which the path response is already sent, PCE accepts the request and process it. This is because the PCE may not be able to identify that the received request is duplicate unless otherwise it maintains the history.

- o When a PCC receives a duplicate response message, it discards it as it will not be able to locate the original request message that was processed earlier.

#### 4.3. Congestion Control

UDP does not provide any means to handle congestion. Moreover, when UDP is used as the transport layer protocol, the rate at which the message is generated and transmitted is based on the application capabilities. This may lead to congestion at PCE when multiple PCCs are sending large number of requests. In such case, the PCE MAY drop the additional requests that are governed by the load attributes such as memory, CPU, etc. The PCE MAY not send any notification to the PCC that the request message is dropped. When the request is dropped, PCC will retransmit the PCECP message based on its retransmission strategies.

However, based on a hysteresis, PCE can notify the corresponding PCC before dropping the received response message during overload. At once PCE is recovered from the high load condition it can continue to provide the path computation service.

In the event of congestion in the network, UDP datagrams (carrying PCECP messages) can be dropped and PCC or PCE may be unaware of this. Handling of such messages is not required and is out of scope from this document. However, this results in PCC retransmitting the corresponding request message.

#### 4.4. PCECP Messages and Object Formats

No new messages and objects are introduced in this document. The PCECP messages and objects are the same as defined in [[RFC5440](#)].

### 5. UDP Port

The PCECP operates over UDP using a registered UDP port (4189). All PCECP messages MUST be sent using the registered UDP port.

## 6. Security Considerations

The PCECP messages could be the target of the following security issues:

- o Spoofing (PCC or PCE impersonation)
- o Snooping (message interception)
- o Falsification
- o Denial of Service

When UDP is used, the application MUST address these security issues. This can be achieved using authenticated message exchange along with an appropriate replay detection scheme. At once authentication is enabled; PCECP peers can verify the integrity of message before processing it. PCECP peer drops the received message in the event of authentication failure.

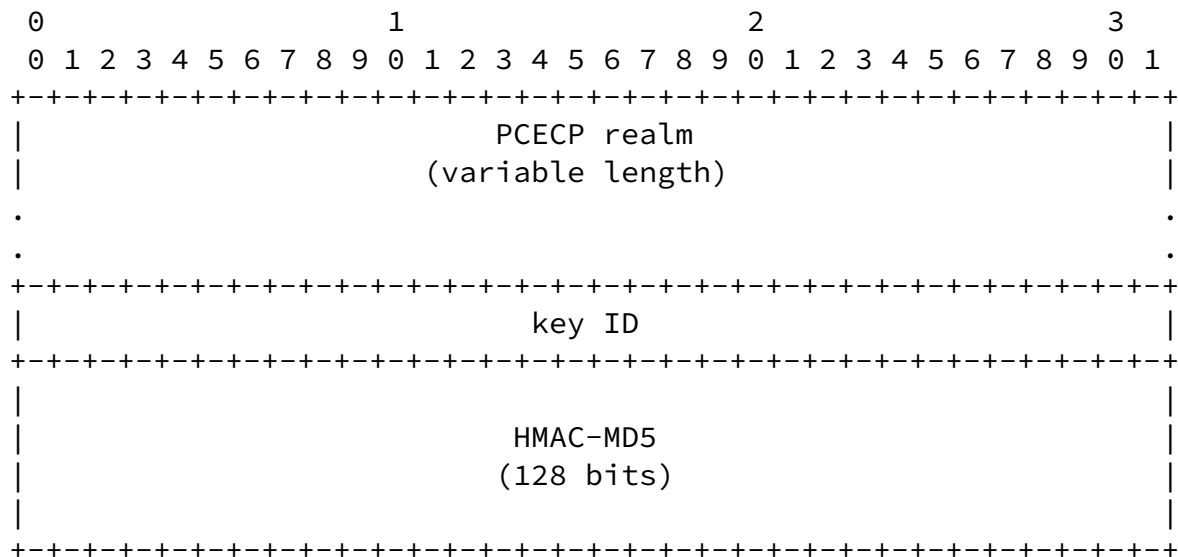
### 6.1. Authentication of PCECP Messages

The authentication of PCECP messages plays a vital role in resolving the falsification of messages, integrity, and denial of service attacks. PCECP peers can be preconfigured with authentication keys to be used during the message exchange. The authentication keys can be configured manually or by employing an appropriate key-exchange protocol which is out of scope from this document.

The authentication of PCECP message can be achieved by carrying the authentication information in the Authentication object to reliably identify the source of a PCECP message and to confirm that the contents of the PCECP message have not been tampered with.

The Authentication object carries authentication information to authenticate the identity and contents of PCECP messages. The format of the Authentication object is as follows:





PCECP realm: The PCECP realm identifies the administrative domain under which PCC and PCE are deployed.

key ID: The key identifier that identifies the key used to generate the HMAC-MD5 value.

HMAC-MD5: The message authentication code generated by applying MD5 to the PCECP message using the key identified by the PCECP realm, PCC IP address, and key ID.

The sender computes the MAC using the HMAC generation algorithm and the MD5 hash function. The entire PCECP message (setting the MAC field of the authentication option to zero), including the PCECP message header and the options field, is used as input to the HMAC-MD5 computation function.

### 6.1.3. Message Validation

To validate an incoming message, the receiver computes the MAC. The entire PCECP message (setting the MAC field of the authentication option to 0) is used as input to the HMAC-MD5 computation function. If the MAC computed by the receiver does not match the MAC contained in the authentication option, the receiver **MUST** discard the PCECP message.

---

[draft-zhang-pce-udp-for-pcecp-00.txt](#)

January 2010

#### 6.1.4. Key Utilization

Each PCC and PCE is assigned with a set of key. Each key is uniquely identified by the IP address of the peer.

The PCC and PCE use the assigned keys to authenticate PCECP messages during a session.

## [7. Conclusion](#)

UDP can be used as the transport layer protocol for PCECP instead of TCP. The application will become responsible for the reliable delivery of the messages and also monitors the congestion. The PCECP session holds good even when UDP being the transport layer protocol.

## [8. IANA Considerations](#)

### [8.1. UDP Port](#)

PCECP will use a registered UDP port to be assigned by IANA (4189).

### [8.2. PCECP Objects](#)

The PCECP Objects registry contains a sub registry, PCECP Objects.

IANA is requested to make some allocations for the Authentication object.

Object-Class Value	Name	Reference
27	Authentication Object-Type-1	This document

### [8.3. PCECP TLV Type Indicators](#)

IANA is requested to create a registry for the following TLVs that appear in the Authentication object.

Value	Meaning	Reference
1	RDM Monotonic Counter TLV	This document
2	HMAC-MD5 TLV	This document

---

[draft-zhang-pce-udp-for-pcecp-00.txt](#)

January 2010

## [9.](#) Acknowledgments

The authors would like to thank Adrian Farrel, Pradeep Shastry, Thiyagarajan Manickam, Hemalatha G for their suggestions during the development of this draft.

## [10.](#) References

### [10.1.](#) Normative References

- [RFC1321] Rivet, R., "The MD5 Message-Digest Algorithm", April 1992.
- [RFC2104] Canetti, R., Bellare, M., and H. Krawczyk, "HMAC: Keyed-Hashing for Message Authentication", February 1997.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", March 1997.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", May 2008.
- [RFC768] Postel, J., "User Datagram Protocol", August 1980.

### [10.2.](#) Informative References

- [RFC4655] Vasseur, J. and J. Ash, "A Path Computation Element (PCE)-Based Architecture", August 2006.
- [RFC4657] Ash, J. and J. Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", September 2006.
- [RFC5440] Ash, J. and J. Roux, "Path Computation Element (PCE) communication Protocol (PCEP)", March 2009.

## [11.](#) Authors' Addresses

Fatai Zhang  
Huawei Technologies  
F3-5-B R&D Center, Huawei Base  
Bantian, Longgang District  
Shenzhen 518129 P.R.China

Phone: +86-755-28972912

zhang

Expires July 2010

[Page 14]

---

[draft-zhang-pce-udp-for-pcecp-00.txt](#)

January 2010

Email: zhangfatai@huawei.com

Suresh BR  
Huawei Technologies  
Shenzhen  
China

Email: sureshbr@huawei.com

SenthilKumar S  
Huawei Technologies  
Shenzhen  
China

Email: ssenthilkumar@huawei.com

Jun Sun  
Huawei Technologies  
Shenzhen  
China

Phone: +86-755-28977297  
Email: johnsun@huawei.com

## Intellectual Property

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology



described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>

zhang

Expires July 2010

[Page 15]

---

[draft-zhang-pce-udp-for-pcecp-00.txt](#)

January 2010

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at [ietf-ipr@ietf.org](mailto:ietf-ipr@ietf.org).

The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions.

For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of [RFC 5378](#). No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under [RFC 5378](#), shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

#### Disclaimer of Validity

All IETF Documents and the information contained therein are provided

on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

#### Full Copyright Statement

Copyright (c) 2009 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents in effect on the date of publication of this document (<http://trustee.ietf.org/license-info>).

zhang

Expires July 2010

[Page 16]

---

[draft-zhang-pce-udp-for-pcecp-00.txt](#)

January 2010

Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

